

System and method for noise-compensated speech recognition

Publication number: CN1296607

Publication date: 2001-05-23

Inventor: SIN G C (US); NING BI (US)

Applicant: QUALCOMM INC (US)

Classification:

- **international:** **G10L15/06; G10L15/02; G10L15/10; G10L15/20; G10L15/00;** (IPC1-7): G10L15/20

- **European:** G10L15/20

Application number: CN19998004886 19990203

Priority number(s): US19980018257 19980204

Also published as:

WO9940571 (A1)
EP1058925 (A1)
US6381569 (B1)
US2001001141 (A1)
EP1058925 (A0)

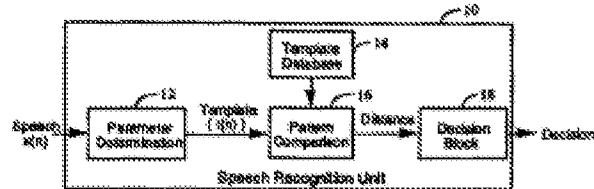
more >>

Report a data error here

Abstract not available for CN1296607

Abstract of corresponding document: **WO9940571**

A system and method for improving speech recognition accuracy in the presence of noise is provided. The speech recognition training unit is modified to store digitized speech samples into a speech database that can be accessed at recognition time. The improved recognition unit comprises a noise analysis, modelling, and synthesis unit which continually analyzes the noise characteristics present in the audio environment and produces an estimated noise signal with similar characteristics. The recognition unit then constructs a noise-compensated template database by adding the estimated noise signal to each of the speech samples in the speech database and performing parameter determination on the resulting sums. This procedure accounts for the presence of noise in the recognition phase by retraining all the templates using an estimated noise signal with similar characteristics as the actual noise signal that corrupted the word to be recognized. This method improves the likelihood of a good template match, which increases the recognition accuracy.



Data supplied from the **esp@cenet** database - Worldwide

[12] 发明专利申请公开说明书

[21] 申请号 99804886.0

[43] 公开日 2001 年 5 月 23 日

[11] 公开号 CN 1296607A

[22] 申请日 1999.2.3 [21] 申请号 99804886.0

[30] 优先权

[32] 1998.2.4 [33] US [31] 09/018,257

[86] 国际申请 PCT/US99/02280 1999.2.3

[87] 国际公布 WO99/40571 英 1999.8.12

[85] 进入国家阶段日期 2000.10.8

[71] 申请人 夸尔柯姆股份有限公司

地址 美国加州圣地埃哥

[72] 发明人 G·C·西

毕 宁

[74] 专利代理机构 上海专利商标事务所

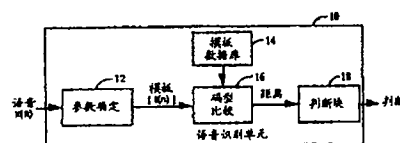
代理人 李家麟

权利要求书 4 页 说明书 7 页 附图页数 2 页

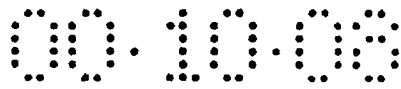
[54] 发明名称 用于经噪声补偿的话音识别的系统和方法

[57] 摘要

一种提高有噪声存在时语音识别准确性的系统和方法。改进语音识别训练单元,将数字化的语音取样存入可以在识别时间里访问的语音数据库。改进的识别单元包含一个噪声分析、模型化和合成单元,它连续地分析声音环境中存在的噪声特征,并产生具有相似特征的估计噪声信号。随后,通过将估计的噪声信号加到语音数据库中的每一语音样本中,并对得到的结果和进行参数确定,识别单元构造一个噪声补偿的模板数据库。这一过程考虑到识别阶段中有噪声存在,采用具有相似特征的估计噪声作为侵扰要识别的字的实际噪声信号,对训练所有的模板。这一方法提高了良好模板匹配的可能性,提高了识别准确性。



ISSN 1008-4274



权 利 要 求 书

1. 一种语音识别系统，其特征在于，它包含：训练单元，用来接收要训练的词语的信号、产生用于每一所述词语的数字化取样，并将所述数字化取样存储在语音数据库中；以及

语音识别单元，用来接收要识别的受噪声侵扰的输入信号、通过将噪声的影响施加于所述语音数据库的所述数字化取样上来产生噪声补偿的模板数据库，并根据所述噪声补偿的模板数据库，提供用于所述受噪声侵扰输入信号的语音识别结果。

2. 如权利要求 1 所述的语音识别系统，其特征在于，所述语音识别单元包含：

第一参数确定单元，用来接收所述受噪声侵扰的输入信号，并按照预定的参数确定技术，产生代表所述输入信号的参数模板；

第二参数确定单元，用来接收所述语音数据库，所述数据库具有作用于所述数字化取样的噪声影响，并按照所述预定的参数确定技术，产生所述噪声补偿的模板数据库；以及

码型比较单元，用来代表所述输入信号的所述参数模板与所述噪声补偿的模板数据库比较，以确定最佳匹配，从而识别所述语音识别结果。

3. 如权利要求 1 所述的语音识别系统，其特征在于，所述语音识别单元包含：

语音检测单元，用来接收所述受噪声侵扰的输入信号，并判断所述输入信号中是否有语音，其中，所述输入信号是在判断所述输入信号中没有语音时指定为噪声信号；以及

噪声单元，所述噪声单元是在判断所述输入信号中没有语音时启动的，所述噪声单元用来分析所述噪声信号并合成具有所述噪声信号特征的合成噪声信号，所述合成噪声信号用来将噪声影响作用于所述语音数据库的所述数字化取样。

4. 如权利要求 2 所述的语音识别系统，其特征在于，所述语音识别单元还包含：

语音检测单元，用来接收受噪声侵扰的输入信号，并判断所述输入信号中是否有语音，其中，当判断在所述输入信号中没有语音时，所述输入信号被指定为是噪声信号；以及

在判断所述输入信号中没有语音时启动的噪声单元，所述噪声单元用来分析所述噪声信号，并合成具有所述噪声信号的特征的合成的噪声信号，所述合成的噪声信号用来向所述语音数据库的所述数字化取样施加噪声影响。

5. 如权利要求 2 所述的语音识别系统，其特征在于，所述参数确定技术是一种线性预告编码(LPC)分析技术。

6. 如权利要求 4 所述的语音识别系统，其特征在于，所述参数确定技术是一种线性预告编码(LPC)分析技术。

7. 如权利要求 3 所述的语音识别系统，其特征在于，所述语音检测单元通过分析所述输入信号中语音活动的电平，来判断是否存在语音。

8. 如权利要求 4 所述的语音识别系统，其特征在于，所述语音检测单元通过分析所述输入信号中语音活动的电平来判断是否存在语音。

9. 如权利要求 3 所述的语音识别系统，其特征在于，所述噪声单元用线性预告编码(LPC)技术，来分析和合成所述合成的噪声信号。

10. 如权利要求 3 所述的语音识别系统，其特征在于，所述合成的噪声信号与要识别的所述输入信号前，记录的所述噪声信号的窗口对应。

11. 如权利要求 3 所述的语音识别系统，其特征在于，所述合成的噪声信号与预定时间内记录的所述噪声信号的各个窗口平均值对应。

12. 如权利要求 4 所述的语音识别系统，其特征在于，所述噪声单元用线性预告编码(LPC)技术分析和合成所述合成的噪声信号。

13. 如权利要求 4 所述的语音识别系统，其特征在于，所述合成的噪声信号与要识别的所述输入信号前记录的所述噪声信号窗口对应。

14. 如权利要求 4 所述的语音识别系统，其特征在于，所述合成的噪声信号与在预定时间内记录的所述噪声信号的各个窗口平均值对应。

15. 一种考虑到噪声环境影响的语音识别系统的训练单元，其特征在于，它包含：

接收要训练的词语信号的装置；

产生用于每一所述词语的数字化取样的装置；以及

将所述数字化取样存入语音数据库的装置。

16. 一种语音识别系统的语音识别单元，用来识别一输入信号，所述语音识别单元考虑到噪声环境的影响，其特征在于，它包含：

将一词汇表的词语数字化取样存入语音数据库中的装置；使噪声影响作用于

声侵扰的语音信号 $r(n)$ 输入到参数确定块 24, 产生受噪声侵扰的模板 $t1(n)$ 。码型比较块 28 将模板 $t1(n)$ 与模板数据库 26 中的所有模板比较, 而模板数据库 26 是构筑在安静环境中的。由于受噪声侵扰的模板 $t1(n)$ 不是与训练模板中的任何一个匹配的, 所以, 有这样一种很高的可能性, 即由判断块 30 产生的判断可能是一个识别错误或失败。

发明概述

本发明是一种在有噪声时对说出的词语进行自动识别的系统和方法。依赖于说话者的语音识别系统分两个阶段工作: 训练阶段和识别阶段。在传统的语音识别系统训练阶段, 提示使用者说出规定的词汇表中的所有的词语。对每一词语的数字化语音模板进行处理, 以产生表征说出的词语的参数模板。训练阶段的输出是一个这些模板组成的信息库。在识别阶段, 使用者说出特定的词语, 来启动想要的动作。使说出的词语数字化, 并进行处理, 产生一个模板, 并将其与所有在训练期间产生的模板比较。最接近的匹配确定将要执行的动作。使语音识别系统的准确性受到限制的主要损害是在有噪声的时候。识别期间加入噪声严重损害了识别的准确性, 这是因为噪声是不在产生模板数据库时的训练期间出现的。本发明认识到, 需要考虑到在识别时出现的特定噪声, 以提高识别的准确性。

因此, 不采用存储参数模板, 改进的语音处理系统和方法存储训练阶段每一说出的词语的语音模板, 并使其数字化。因此, 训练阶段的输出是一个数字化的语音数据库。在识别阶段, 连续监视声音环境中的噪声特征。当使用者说出词语进行识别时, 通过在语音数据库中的每一信号中加入噪声信号并对每一语音加噪声信号进行判断构筑经噪声补偿的模板数据库。该加入的噪声信号的每一实施例是一个人工合成的噪声信号, 具有与实际噪声类似的特征。另一种实施例是记录一个在使用者说出词语进行识别前出现的噪声时间窗。由于模板数据库是用在要识别的词语中出现的相同类型的噪声构筑的, 所以语音识别单元可以找到模板之间的良好匹配, 提高识别准确度。

附图简述

读者在参照附图阅读了本发明的详细描述以后, 将会更清楚地了解本发明的特征、目的和优点。图中, 相同的标号所表示的意义相同。

图 1 是语音识别系统训练单元的方框图;

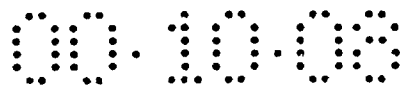


图 2 是语音识别单元的方框图；

图 3 是对受噪声侵扰的语音输入进行语音识别的语音识别单元方框图；

图 4 是改进的语音识别系统训练单元的方框图；以及

图 5 是典型的改进的语音识别单元的方框图。

较佳实施例的详细描述

本发明提供了一种在有噪声时提高语音识别准确性的系统和方法。它利用了近年来计算能力和存储器集成进步的优点，并修改了考虑到在识别时有噪声出现时的训练和识别阶段。语音识别单元的功能是找到与识别模板最接近的匹配，而该模板是对受噪声侵扰的话音进行计算的。由于噪声的特征会随时间和地点而变，本发明认为，构筑模板的最佳时间是在识别阶段。

图 4 给出的是语音识别系统改进的训练单元 40 的方框图。与图 1 中传统的训练方法不同的是，修改了训练单元 40，去掉了参数确定步骤。与存储参数模板不同的是，存储的是实际词语的数字化语音采样。所以，训练单元 40 接收作为输入语音取样 $s(n)$ ，并在语音数据库 42 中存储数字化的语音取样 $s(n)$ 。在训练以后，语音数据库 42 含有 M 个语音信号，这里， M 是词汇表中的词汇数。而已有技术的参数确定系统和方法丢失了有关语音特征的信息，它仅存储语音参数，该系统和方法会保留所有的语音信息，用于识别阶段。

图 5 是与训练单元 40 一起使用的改进的语音识别单元 50 的方框图。语音识别单元 50 的输入是受噪声侵扰的语音信号 $r(n)$ 。受噪声侵扰的语音信号 $r(n)$ 是由加法器 52 通过把语音信号 $x(n)$ 与噪声信号 $w(n)$ 相加后得到的。与以前一样，加法器 52 不是系统的一个具体元件，但是噪声环境的模拟。

语音识别单元 50 包含语音数据库 60，它含有数字化的语音取样，这些取样是在训练阶段记录的。语音识别单元 50 还包含参数确定块 54，通过它传送受噪声侵扰的语音信号 $r(n)$ ，而产生受噪声侵扰的模板 $t1(n)$ 。正如在传统的话音识别系统中，参数确定块 54 采用任何一种语音参数确定技术都可以。

典型的参数确定技术采用线性预告编码(LPC)分析技术。LPC 分析技术模拟声道(vocal tract)作为数字滤波器。采用 LPC 分析，LPC 倒频谱系数 $c(m)$ 可以计算为参数，用以代表语音信号。系数 $c(m)$ 是用下述步骤计算的。首先，通过采用窗口函数 $v(n)$ ，对一个语音取样帧，对受噪声侵扰的语音信号 $r(n)$ 开视窗：

$$y(n) = r(n)v(n) \quad 0 \leq n \leq N-1 \quad (1)$$

在本典型实施例中，窗口函数 $v(n)$ 是一个汉明窗口，并且帧大小 N 等于 160。接着，用下面的等式，对窗口取样进行自相关系数计算：

$$R(k) = \sum_{m=0}^{N-k} y(m)y(m+k) \quad K=1, 2, \dots, P \quad (2)$$

在典型的实施例中， P 是要计算的自相关系数，等于 LPC 预测器的阶数，它等于 10。随后，用 Durbin 递归规则，直接从自相关值计算 LPC 系数。该规则可以陈述如下：

$$1. E^{(0)} = R(0), \quad i=1 \quad (3)$$

$$2. k_i = \left\{ R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right\} / E^{(i-1)} \quad (4)$$

$$3. \alpha_i^{(i)} = k_i \quad (5)$$

$$4. \alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \leq j \leq i-1 \quad (6)$$

$$5. E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (7)$$

$$6. \text{如果 } i < P, \text{ 则回到 (2), 并且 } i=i+1 \quad (8)$$

7. LPC 系数的最终解给出

$$a_j = \alpha_j^{(P)} \quad 1 \leq j \leq P \quad (9)$$

随后，用下面的等式，将 LPC 系数转换成 LPC 倒频谱系数：

$$c(0) = \ln(R(0)) \quad (10)$$

$$c(m) = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k} \quad 1 \leq m \leq P \quad (11)$$

$$c(m) = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k} \quad m > P \quad (12)$$

应当理解，其他技术可以用于参数确定，而取代 LPC 倒频谱系数。

另外，将信号 $R(n)$ 传送到语音检测块 56，由它判断有无语音。语音检测块 56 可以用任意一种技术来判断是否有语音存在。这样的一种方法见上述美国专利 5,414,796，其标题是“可变速率声音合成机”。这一技术对话音活动的电平进行分析，对有无语音存在作出判断。语音活动电平是基于与背景噪声能量估计的比较的信号能量的。首先，对每一帧计算能量 $E(n)$ ，在一较佳实施例中，是由 160

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (21)$$

这是用于噪声分析的滤波器的逆。在对每一合成噪声取样取一换算倍数形成与实际噪声能量相等的合成噪声能量以后，输出为合成噪声 $w1(n)$ 。

合成的噪声 $w1(n)$ 由加法器 62 加到语音数据库 60 中的每一组数字化语音取样中，得到合成的受噪声侵扰的语音取样集。随后，每一组合成的受噪声侵扰的语音取样通过参数确定块 64，与参数确定块 54 中使用的一样，用相同的参数确定技术，产生用于每一组合成的受噪声侵扰的语音取样的一组参数。参数确定块 54 产生用于每一组语音取样的参数模板，并且将这些模板存储在噪声受到补偿的模板数据库 66 中。噪声补偿的模板数据库 66 是一组这样构筑的模板，就象是在识别期间出现的相同类型的噪声已经发生了传统的训练。注意，除了美国专利 5,646,991 中揭示的方法以外，有许多可能的方法，用来产生估计噪声 $w1(n)$ 。另一种实施例是简单记录实际当使用者沉默时出现的实际噪声的时间窗口，并用该噪声信号作为估计噪声 $w1(n)$ 。在说出要识别的词语前所记录的噪声时间窗口是本方法的典型实施例。另一种方法是对在指定时间周期内得到的各个噪声窗口取平均。

再参见图 5，码型比较块 68 将受噪声侵扰的模板 $t1(n)$ 与噪声受补偿的模板数据库 66 中的所有模板比较。由于噪声的影响包括在噪声受到补偿的模板数据库 66 的模板内的，所以，判断块 70 能够为 $t1(n)$ 找到很好的匹配。考虑到以这种方式噪声的影响，可以提高语音识别系统的准确性。

上述较佳实施例的描述使得本领域的技术人员能够制作和使用本发明。很明显，本领域中的技术人员还能够对这些实施例作各种修改，并且在无需发明专业人员的帮助下，还可以将这些基本原理应用于其他的实施例。所以，本发明并非仅限于所给的实施例，应当从最宽的范围来理解本发明的原理和新特征。

个取样组成的。随后，用下面的等式计算背景噪声能量估计 $B(n)$ ：

$$B(n) = \min[E(n), 5059644, \max(1.00547 * B(n-1), B(n-1) + 1)] \quad (13)$$

如果 $B(n) < 160000$ ，则如下所述，用 $B(n)$ 计算三个阈值：

$$T1(B(n)) = -(5.544613 \times 10^{-6}) * B^2(n) + 4.047152 * B(n) + 362 \quad (14)$$

$$T2(B(n)) = -(1.529733 \times 10^{-5}) * B^2(n) + 8.750045 * B(n) + 1136 \quad (15)$$

$$T3(B(n)) = -(3.957050 \times 10^{-5}) * B^2(n) + 18.89962 * B(n) + 3347 \quad (16)$$

如果 $B(n) > 160000$ ，则三个阈值计算为：

$$T1(B(n)) = -(9.043945 \times 10^{-8}) * B^2(n) + 3.535748 * B(n) - 62071 \quad (17)$$

$$T2(B(n)) = -(1.986007 \times 10^{-7}) * B^2(n) + 4.941658 * B(n) + 223951 \quad (18)$$

$$T3(B(n)) = -(4.838477 \times 10^{-7}) * B^2(n) + 8.630020 * B(n) + 645864 \quad (19)$$

该语音检测方法表示，当能量 $E(n)$ 大于阈值 $T2(B(n))$ 时有语音存在，并且表示当能量 $E(n)$ 小于阈值 $T2(B(n))$ 时，没有语音存在。在另一种实施例中，这一方法可以扩展到在两个或多个频带中计算背景噪声能量估计和阈值。另外，应当理解，等式(13)－(19)中的数值是通过实验确定的，并且可以根据环境来修改。

当语音检测块 56 判断不存在语音时，它发送一个使得能够启动噪声分析、建模和合成块 58 的控制信号。应当注意，在没有语音时，接收信号 $R(n)$ 与噪声信号 $w(n)$ 是相同的。

当启动噪声分析、建模和合成块 58 时，它分析噪声信号 $r(n)$ 的特征，对其进行建模，并合成一个具有与实际噪声 $w(n)$ 相似的特征的噪声信号 $w1(n)$ 。进行噪声分析、建模和合成的典型实施例见美国专利 5,646,991，其标题是“回波消除器中的噪声取代系统和方法(Noise Replacement System and Method in an Echo Canceller)”，该专利已转让给本发明的受让人，在此引述供参考。该方法通过预计差错滤波器传送噪声信号 $r(n)$ 来进行噪声分析：

$$A(z) = 1 - \sum_{i=1}^P a_i z^{-i} \quad (20)$$

这里， P 是预测器的阶数(order)，在本典型实施例中是 5。LPC 系数 a_i 是用等式(1)到(9)，如先前所说明的那样计算的。一旦得到了 LPC 系数，则通过噪声合成滤波器传送白噪声，可以产生合成噪声取样，它们具有相同的频谱特征，如下式所述：

说明书附图

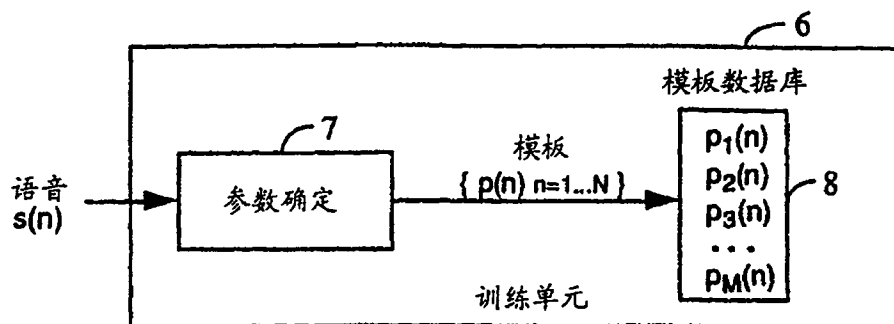


图 1

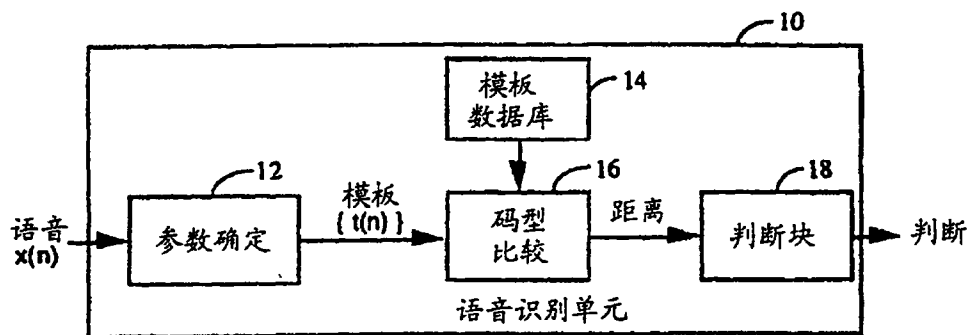


图 2

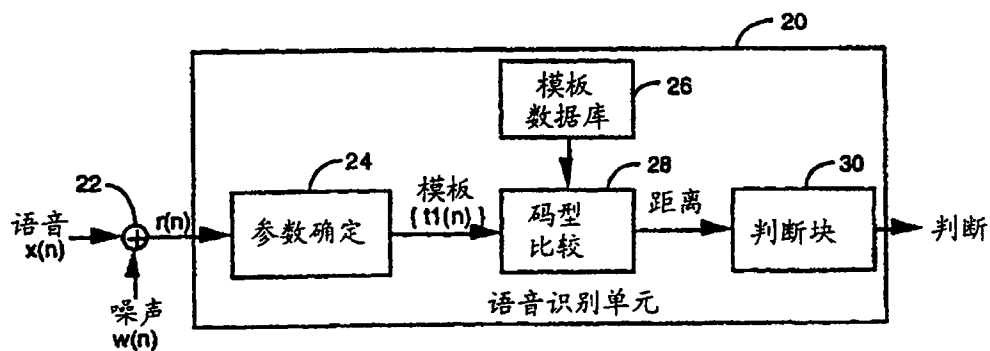


图 3

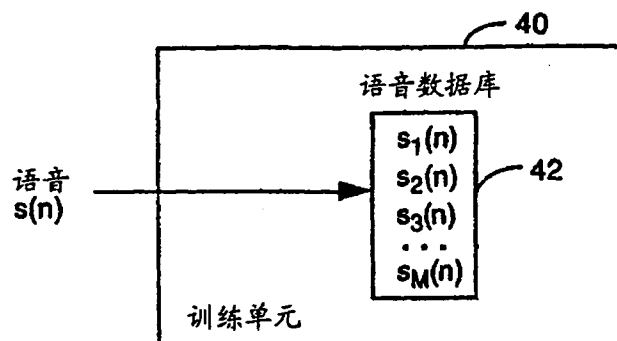


图 4

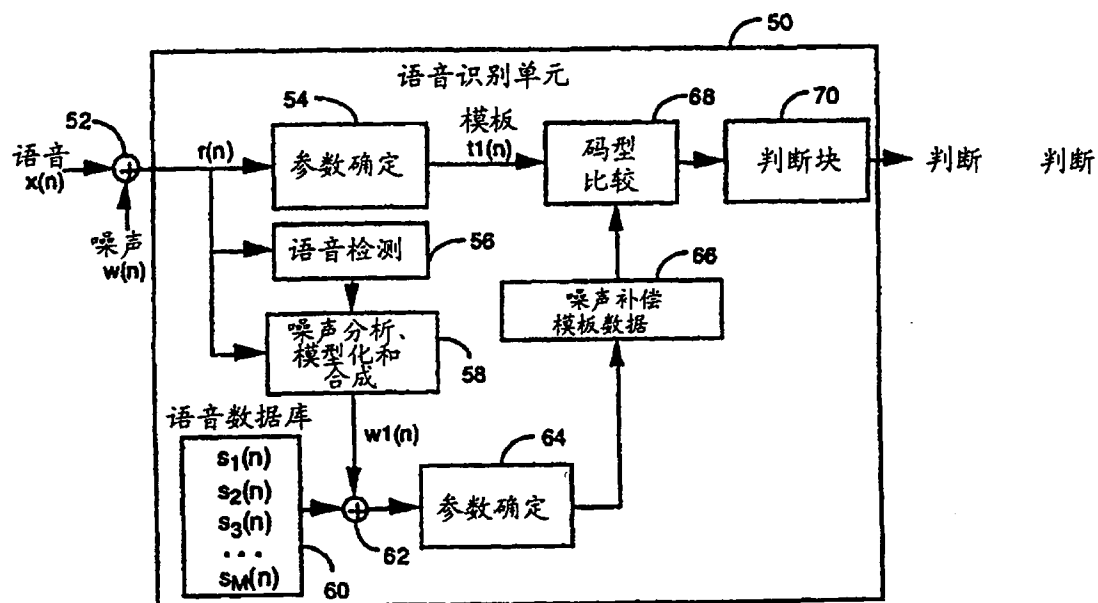


图 5